

The Gradient of Generative AI Release: Methods and Considerations

Year: 2023 | Citations: 122 | Authors: Irene Solaiman

Abstract

As increasingly powerful generative AI systems are developed, the release method greatly varies. We propose a framework to assess six levels of access to generative AI systems: fully closed; gradual or staged access; hosted access; cloud-based or API access; downloadable access; and fully open. Each level, from fully closed to fully open, can be viewed as an option along a gradient. We outline key considerations across this gradient: release methods come with tradeoffs, especially around the tension between concentrating power and mitigating risks. Diverse and multidisciplinary perspectives are needed to examine and mitigate risk in generative AI systems from conception to deployment. We show trends in generative system release over time, noting closedness among large companies for powerful systems and openness among organizations founded on principles of openness. We also enumerate safety controls and guardrails for generative systems and necessary investments to improve future releases.